Lecture 1 September 22nd, 2025

1. Events, probabilities, and Bayes rule
2. Bayesian inference for single variable models

Review of probability concepts

Random experiment/process: an experiment or process, where deterministic prediction is hard

Example: coin toss

When we toss a coin with high initial velocity it is hard to predict how the coin will land.

Random experiments generate simple events

Example: a random experiment of tossing a coin twice results in the following simple events:

HH, HT, TH, TT, where H = heads, T = tails

A set of <u>all</u> possible simple events is called a sample space. For the coin tossing experiment above the sample space is a set

$$S = \{HH, HT, TH, TT\}$$

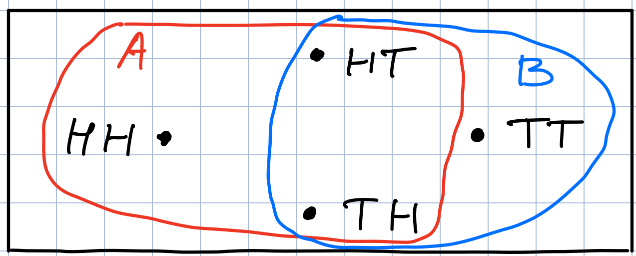
An event (not necessarily simple) is any subset of the sample space.

Example: In coin tossing experiment we can define the following events:

$A =$ "at least one H" $\Rightarrow A = \{HH, HT, TH\}$

$B =$ "at most one H" $\Rightarrow B = \{HT, TH, TT\}$

Since events are sets it is sometimes useful to visualize them with Venn Diagrams:



If events cannot occur at the same time they are called mutually exclusive (disjoint in terms of set theory)

Event/set arithmetic

$A \overset{\text{or}}{\cup} B$ = { simple events in A or B, or both }

$A \overset{\text{and}}{\cap} B$ = { simple events in both A and B }

$\overset{\text{complement}}{A^c}$ = { simple events not in A }

Example: twice tossing a coin

$A =$ "at least one H" $= \{HH, HT, TH\}$

$B =$ "at most one H" $= \{HT, TH, TT\}$

$A \cup B = \{HH, HT, TH, TT\} = S$ - the whole sample space

$A \cap B = \{HT, TH\}$

$A^c = \{TT\}$, $B^c = \{HH\}$

Probability axioms:

subset   probability of event A

1) For any event $A \subset S$, $0 \le P(A) \le 1$
2) $P(S) = 1$
3) For mutually exclusive events A and B $(A \cap B = \varnothing)$, $P(A \cup B) = P(A) + P(B)$

Conditional probability

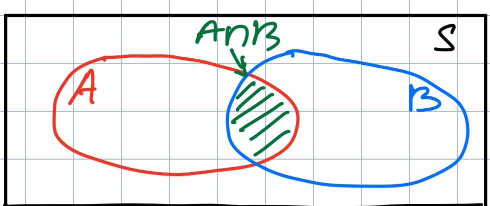
Sometimes we want probability of event A knowing that event B has occurred. Intuitively,

P("being struck by lightning") < P("being struck by lightning | "caught in a storm")

conditioning sign

Conditional probability is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$, where $P(B) > 0$ because it doesn't make sense to condition on improbable events



Inverting Conditional Statements

$$P(A \cap B) = P(A \mid B)\,P(B); \quad \boxed{P(A \cap B) = P(B \mid A)\,P(A)}$$

Bayes rule: $$\boxed{P(A \mid B)} = \frac{P(A \cap B)}{P(B)} = \frac{\boxed{P(B \mid A)}\,P(A)}{P(B)}$$

Key players: $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \mid B)$, $P(B \mid A)$

Example (medical testing)

Suppose we are evaluating a medical test for a disease D. No test is perfect. This particular test has false-negative rate of 1% and false positive rate of 2% . We also know the frequency of the disease in the population is 0.1%.

Question: what is the probability that a patient has the disease if their test was positive?

Let's define events of interest:

$T+$ = "test is positive" $\quad D+$ = "disease present"

$T-$ = "test is negative" $\quad D-$ = "disease absent"

What probabilities are given to us?

$$P(T+|D+) = 0.99 \qquad P(T+|D-) = 0.02$$

$$P(T-|D+) = 0.01 \qquad P(T-|D-) = 0.98$$

$$P(D+) = 0.001$$

What probability do we need to compute?

$$P(D+|T+) = \frac{P(D+\cap T+)}{P(T+)} = \frac{P(T+|D+)\,P(D+)}{P(T+)}$$

$P(T+) = P((T+\cap D+)\cup(T+\cap D-))$ = [$T+\cap D+$ and $T+\cap D-$ are mutually exclusive] = $P(T+\cap D+) + P(T+\cap D-)$ =

= [inverting conditional probability definition] =

$= P(T+|D+)P(D+) + P(T+|D-) \cdot P(D-) = 0.99 \cdot 0.001$
$+ 0.02 \cdot (1 - 0.001) \approx 0.02$

Back to our calculation:

$$P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+)} = \frac{0.99 \cdot 0.001}{0.02} \approx 0.05 -$$

small number

Let's redo the calculation with $P(D+) = 0.01$
everything else will be the same:

$$P(T+) = 0.99 \cdot 0.01 + 0.02(1 - 0.01) \approx 0.03$$

$$P(D+|T+) = \frac{0.99 \cdot 0.01}{0.03} = 0.33$$ – much larger number

Note: we have just performed our first Bayesian inference: we started from prior knowledge ($P(D+)$), updated this prior with data ($T+$ is true) and arrived at the posterior probability $P(D+|T+)$.

Bayesian inference for single parameter

So far we've seen only discrete objects (events). To learn/recover continuous objects/parameters from data, we need to recall what random variables are:

A random variable is a function mapping sample space S into $\mathbb{R}^n$ – n-dimensional space of real numbers.

Example : coin flipping

Sample space

$S = \{HH, HT, TH, TT\}$

Random variable : # of heads
$(g : S \to \mathbb{R})$

simple event	g (simple event)	
HH	2	0.25
HT	1	0.25
TH	1	0.25
TT	0	0.25

g	2	1	0
P(g=x)	0.25	0.5	0.25

Disrete random variables

prob mass ftn: (pmf)

k	0	1	2	3	...	n
p(k)	p(0)	p(1)	p(2)	p(3)		p(n)

$\sum_{k=0}^{n} p(k) = 1$

$p(k) = P(X = k)$

Bernoulli $X \sim \text{Bernoulli}(p)$.

X	0	1
Pr	1-p	p

pmf: $p(0) = 1-p$, $p(1) = p$, $E(X) = p$, $Var(X) = p(1-p)$

Binomial $Y \sim \text{Binomial}(n, p)$. $Y = \sum_{i=1}^{n} X_i$, $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$

pmf: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$, $E(X) = np$, $Var(X) = np(1-p)$

X	0	1	...	n
Pr				

Cumulative distribution function (cdf)

$F(x) = P(X \le x)$ for all $x \in \mathbb{R}$

Continuous random variables

prob density ftn: (pdf) $f(x) \ge 0$ such that $\underbrace{F(x)}_{\text{cdf}} = \int_{-\infty}^{x} \underbrace{f(y)}_{\text{pdf}}\, dy$

Beta $X \sim \text{Beta}(\alpha, \beta)$, $0 < x < 1$, $\alpha > 0$, $\beta > 0$

pdf: $$f(x) = \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{normalizing constant, doesn't depend on } x} x^{\alpha-1}(1-x)^{\beta-1}$$

We are now ready to define our first Bayesian model / problem:

Suppose we would like to estimate the fraction of people in Los Angeles who have been infected (and hopefully recovered) with SARS-CoV-2 virus that causes COVID-19. This can be done by testing for presence of antibodies that we develop in response to the infection. Testing for antibodies the whole population of Los Angeles is impractical. Even testing a large number of people is too expensive, so researchers rectruited 100 random individuals and found that y individuals.

Data: y = number of individuals who had antibodies

$$y \sim \text{Binomial}(100, \theta)$$

came from studies done in other cities

Prior belief: $0.05 < \theta < 0.2$, we can code this prior belief using a Beta distribution:

$$\theta \sim \text{Beta}(\alpha = 2, \beta = 20)$$

Posterior belief: $$\boxed{p(\theta \mid \overset{Y=y}{y})} = \frac{p(\theta, y)}{p(y)} = \frac{\overbrace{p(y\mid\theta)}^{\text{likelihood}}\;\overbrace{p(\theta)}^{\text{prior}}}{\int_0^1 p(\theta, y)\, d\theta} =$$

$$= \frac{\binom{100}{y}\theta^y(1-\theta)^{100-y}\,\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \text{\_\_\_\_\_\_ '' \_\_\_\_\_\_}\, d\theta} =$$

$$= \frac{\cancel{\binom{100}{y}} \cancel{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \, \theta^{y+\alpha-1} (1-\theta)^{100-y+\beta-1}}{\cancel{\binom{100}{y}} \cancel{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \int \theta^{y+\alpha-1} (1-\theta)^{100-y+\beta-1} \, d\theta} = .$$

$$= \frac{\frac{\Gamma(y+\alpha)\,\Gamma(100-y+\beta)}{\Gamma(100+\alpha+\beta)} \, \theta^{y+\alpha-1} (1-\theta)^{100-y+\beta-1}}{\underbrace{\int \frac{\Gamma(y+\alpha)\,\Gamma(100-y+\beta)}{\Gamma(100+\alpha+\beta)} \, \theta^{y+\alpha-1} (1-\theta)^{100-y+\beta-1} \, d\theta}_{=1}} \Rightarrow \theta \mid y \sim Beta(y+\alpha, 100-y+\beta)$$

So if $y=0$, then our posterior belief is

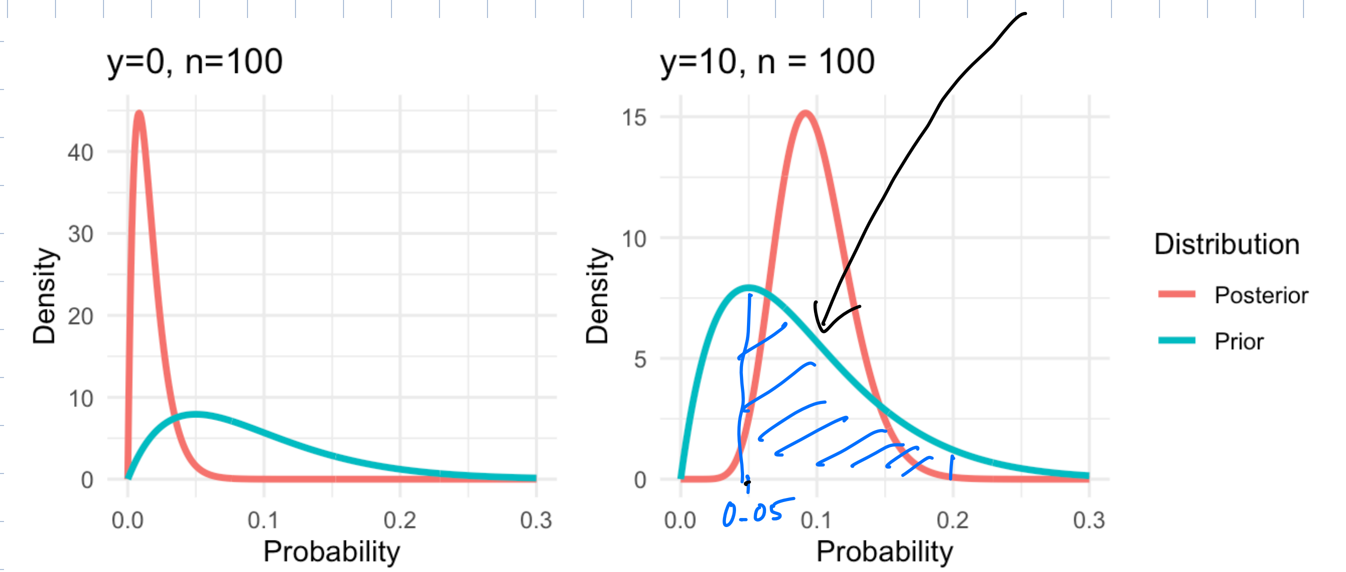
$$\theta \mid y \sim Beta(2, 120)$$

if $y = 10$, then our posterior is

$$\theta \mid y \sim Beta(12, 110)$$

Let's look at our prior and updated posterior beliefs visually:

$$P(0.05 < \theta < 0.2) = 0.66$$



Play with the code if time permits.